BD

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12Q 1/68, C07H 15/12
C07K 3/00, 13/00, A61K 35/14
A01N 63/02, A61K 35/12

A1

(11) International Publication Number: WO 92/12262

(43) International Publication Date: 23 July 1992 (23.07.92)

(21) International Application Number: PCT/US92/00020

(22) International Filing Date: 3 January 1992 (03.01.92)

(30) Priority data:

638,518	4 January 1991 (04.01.91)	US
672,232	20 March 1991 (20.03.91)	US
802,650	6 December 1991 (06.12.91)	US

(71) Applicants: WASHINGTON UNIVERSITY [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). ADELAIDE MEDICAL CENTER FOR WOMEN AND CHILDREN [AU/AU]; 72 King William Road, North Adelaide, S.A. 5006 (AU).

(72) Inventors: SUTHERLAND, Grant, R. ; 21 Thornber Street, Unley Park, S.A. 5061 (AU). RICHARDS, Robert, I. ; 228 Broughm Place, N. Adelaide, S.A. 5006 (AU). SCHLESSINGER, David ; 7475 Teasdale Avenue, St. Louis, MO 63130 (US). NAGARAJA, Ramaiah ; 220 Townsend Lane, St. Louis, MO 63141 (US).

(74) Agent: SWECKER, Robert, S.; Burns, Doane, Swecker & Mathis, George Mason Building, Washington and Prince Streets, P.O. Box 1404, Alexandria, VA 22313-1404 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.
With amended claims.

(54) Title: DNA SEQUENCES RELATED TO ISOLATED FRAGILE X SYNDROME

(57) Abstract

The DNA sequence spanning the fragile X site on the X human chromosome has been obtained in purified and isolated form. As fragile X is associated with mental retardation, the availability of a DNA which spans this locus permits diagnosis and treatment of the related mental disorders.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	RU	Russian Federation
CG	Congo	KP	Democratic People's Republic of Korea	SD	Sudan
CH	Switzerland			SE	Sweden
CI	Côte d'Ivoire	KR	Republic of Korea	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
DE	Germany	LU	Luxembourg	TG	Togo
DK	Denmark	MC	Monaco	US	United States of America

DNA SEQUENCES RELATED TO ISOLATED FRAGILE X SYNDROME

Technical Field

The invention relates generally to genetic diagnosis of humans. In particular, the invention concerns identification of individuals having particular DNA sequences predictive for Fragile X Syndrome.

Background Art

Fragile X syndrome is the most common form of familial mental retardation and affects about one in 2,500 children. The syndrome is characterized by the presence of a cytogenetically detectable fragile site in band q27.3 near the end of the long arm of the X chromosome which, if not the cause of the disorder, is closely associated with it. The diagnostic molecular genetics of the Fragile X Syndrome has been reviewed by Sutherland, G.R. et al. (Clinical Genet. (1990) 37:2-11). An additional review is found by Nussbaum, R.L. et al. (Ann. Rev. Genet. (1986) 20:109-145).

Identification of the DNA spanning and including the fragile site has been reported by Kremer et al. (Am. J. Human Genetics (1991) 49:656-661) and Heitz, et al. (Science (1991) 251: 1236). Characterization of the fragile site has indicated a particular region of

instability within a 5.0 KB EcoRI restriction fragment, with the instability segregating with the Fragile X genotype (Yu et al., Science (1991) 252:1179). The region of instability has further been localized to a 1 KB Pst I fragment containing a $p(CCG)_n$ repeat. The Fragile X genotype is characterized by an increased amount of unstable DNA that maps to the repeat (Kremer et al., Science (1991) 252:1711). The availability of the cloned DNA makes possible the use of the DNA as a probe to detect length polymorphism of the $p(CCG)_n$ to characterize the genotype of an individual at that locus (Kremer et al., supra), thereby obviating problems with cytogenetic visualization at the fragile site (Webb et al., Prenatal Diagnosis (1989) 9:771-781).

Additional diagnostic tools are available in the form of polymorphic microsatellite markers linked to the fragile site at Xq27.3 (FRAXA). Richards, et al., (Am. J. Hum. Genet. (1991) 48:1051-1057) have described polymorphisms associated with length variation in dinucleotide microsatellite repeats in the vicinity of Xq27.3. These markers have a recombination frequency of 1% and 7%, respectively, in two-point linkage analysis in 31 Fragile X families.

Thus, the availability of cloned DNA spanning the fragile site provides reagents uniquely suited for the detection of the Fragile X allele in appropriate subjects. Furthermore, techniques of gene therapy could be used to replace or compensate for the pathologic Fragile X sequence in affected cell types.

Disclosure of the Invention

The invention provides a human DNA sequence corresponding to the Fragile X locus and provides a source for suitable probes for diagnosis and sequences useful for modification in therapy. The obtention of

this sequence from the fragile site thus permits an improvement in diagnostic techniques as well as the possibility for genetic manipulation to overcome the disorder.

In one aspect, the invention is directed to an isolated and purified DNA molecule of no more than 275 kb which includes the fragile site. In another aspect, the invention is directed to a subsequence contained in this larger DNA of no more than 150 kb, which includes the fragile site. In still another aspect, the invention is directed to a DNA probe which crosses the fragile site, and to the corresponding normal sequence useful in replacement therapy.

In still other aspects, the invention is directed to methods to determine the presence or absence of the Fragile X allele in a subject which method comprises probing DNA isolated from the subject with the probe of the invention. Affected individuals appear to have an amplification of a $(CCG)_n$ repeat sequence at the fragile site which gives a band of different size than a normal individual when Southern blots are probed with the probe of the invention.

In another aspect, the invention is directed to oligonucleotides useful as primers in the polymerase chain reaction amplification of polymorphic microsatellite AC repeats closely linked to the Fragile X locus. Thus, these primers may be used to identify alleles of the microsatellite regions which vary in AC repeat length, thereby providing a method for screening for a microsatellite repeat sequence allele predictive of inheritance of the Fragile X allele.

In still another aspect, the invention is directed to methods to correct the fragile site by substituting the normal DNA contained in this region or otherwise compensating for this defect, such as by

administration of the normal protein product or by antibodies directed against the protein product.

Brief Description of the Drawings

Figure 1 shows a schematic of XTY26, a 280 kb plasmid derived from a yeast artificial chromosome (YAC) including a 275 kb human DNA sequence spanning the fragile site.

Figure 2 is a diagram depicting the steps taken in localizing the DNA sequences which comprise the fragile site and the variable region.

Figure 3 depicts the Southern blot analysis of EcoRI-digested somatic hybrid cell DNAs with subclone λ-5, which comprises the 5 kb EcoRI fragment from XTY26.

Figure 4 depicts the Southern blot analysis of EcoRI-digested DNA from two normal and four unrelated Fragile X Syndrome affected males with subclone λ-5.

Figure 5 depicts the Southern blot analysis of PstI-digested DNA from members of the illustrated Fragile X Syndrome pedigree.

Figure 6 illustrates the DNA sequence of the 1.0 kb PstI fragment from a Fragile X Syndrome library.

Figure 7 illustrates the location of primer sequences and polymorphic microsatellite regions for FRAXAc1, FRAXAc2, and FRAXAc3.

Figure 8 illustrates the location of subclones of the PstI region of the fragile site.

Modes of Carrying Out the Invention

Definitions

As used herein, "fragile site" refers to a DNA sequence which occurs at the Xq27.3 locus on the X chromosome in individuals subject to familial mental retardation associated with Fragile X syndrome. "Fragile

X locus" refers to this location whether in normal individuals or in persons affected by the condition.

As used herein, "expression" of fragile X DNA refers to cytogenetic or microscopic manifestation of the fragile site.

All DNA sequences disclosed herein are intended to include complementary sequences, unless otherwise indicated. All DNA sequences are written in a 5'-to-3' direction and conform to nucleotide symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission.

<u>Construction of XTY26</u>

A DNA library was constructed from a human subject known to contain the Fragile X locus by the procedure of Reithman H.C. et al. (<u>Proc. Natl. Acad. Sci. USA</u> (1989) 86:6240). The procedure is designed to rescue telomeres by complementation and was modified by digesting the vector pTYAC1, which propagates in yeast as an "artificial yeast chromosome" (YAC) with BamHI and EcoRI or ClaI to accommodate inserts digested with either EcoRI or TaqI and obtained from the human genomic DNA described below. This method of construction of the YAC library selects for clones which acquire or no longer need an additional telomere. A few clones contain true telomere sequences, others contain segments from nontelomeric regions. Circular chromosomes which are maintained as such in yeast also satisfy the selection (Hieter, P., et al., <u>Cell</u> (1985) 40:381).

The immediate source of the genomic DNA that was inserted in the vector was the human/hamster somatic cell hybrid X3000.11, described by Nussbaum, R.L. et al. (<u>Ann. Rev. Genet</u>. (1986) 20:109-145) which is known to contain a region of human X chromosome from band q24 to qter which spans Xq27.3 and which is known to have the abnormal Fragile X from the original human subject. This

portion of the X chromosome is translocated onto a hamster chromosome in the somatic cell hybrid. The DNA from X3000.11 was digested with TaqI, ligated into pTYACI and transformed into yeast on selective media. The resulting library was screened with the pVK16BI probe known to map close to the fragile site as described by Abidi, F.E., et al. (Genomics (1990) 7:363), and only one clone, XTY26, was positive.

Analysis of the XTY26 clone led to the conclusion that it is a circular YAC with the map shown in Figure 1. *In situ* hybridization was used to determine that the XTY26 clone spanned the fragile site. Total DNA was extracted from yeast cultures containing XTY26 and labeled with fluorescence using the technique of Kievits, T. et al. (Cytogenet. Cell Genet. (1990) 53:134), and the labeled DNA was used as a probe for *in situ* hybridization to metaphase chromosomes expressing Fragile X. The location of the fluorescence labeling relative to the cytogenetically observable fragile site was observed as shown in Table 1. The location of the fluorescence on the chromosome was scored as "proximal," "central" or "distal."

Table 1

Location of Signal for Various Probes in Relation to the Fragile Site at Xq27.3

	Position of signal in relation to fragile site			
Probe	Proximal	Central	Distal	Proximal and Distal
XTY26	11	10	39	8
VK16	10	2	0	0
2-34	9	3	0	0
Do33	0	0	10	0

Sequential metaphase spreads from two Fragile X males were examined until at least ten X chromosomes expressing the fragile site and exhibiting signal from probe hybridization had been scored. The position of the signal was scored as proximal, central (i.e., overlying the gap in the chromosome) or distal to the fragile site. Sensitivity and specificity was such that 35-90% of all metaphases (depending upon the probe) had yellow fluorescent dots on the end of at least one chromatid of the X chromosome with virtually no background signal.

In this context, proximal with respect to the fragile X site means closer to the centromere; distal with respect to fragile X refers to a location closer to the telomere. The majority of the signal was found distal to the fragile site, even though the probe VK16 used to isolate XTY26 was proximal with respect to the Fragile X locus in *in situ* hybridization. The finding of label over proximal, central and distal sites as shown in Table 1 indicates that the clone XTY26 contains DNA complementary to areas of DNA throughout the fragile region.

Additional flanking DNA markers known to map close to the fragile site, Do33 (DXS465) and 2-34 (DXS463), described by Rousseau, F. et al. (*Am. J. Hum. Genet.* (1991) 48:108-116) were also found to be present in XTY26 and their maps for the restriction enzymes BamHI, HindIII and TaqI were identical in both XTY26 and human chromosomal DNA. Because the marker Do33 binds to DNA distal with respect to the fragile site, and marker 2-34 binds to DNA proximal with respect to the fragile site in *in situ* hybridization, their presence in the XTY26 clone supports the conclusion that the DNA insert in the clone spans the fragile region.

The circularity of XTY26 was verified using restriction analysis, and rests on at least four

observations. 1) SalI cuts XTY26 only once and maps within DXS293 which, according to other digests with NaeI, maps toward the middle of the human DNA sequence. The SalI digest gives only a minimal alteration in the size of XTY26 as compared to undigested DNA, consistent with the slight difference between circular and linear DNA of the same mass. 2) DXS293 mapped into the same NruI fragment as 2-34 (140 kb) but to a 120 kb SfiI fragment that was different from the 160 kb SfiI fragment bearing 2-34. The two SfiI fragments (DXS293, 120 kb and 2-34, 160 kb) equalled the total length of XTY26. 3) In addition, 2-34 mapped to within 60 kb of one end of the human DNA insert on an NaeI digest and also to a 50 kb ClaI fragment, yet vector sequences which map to the same 60 kb NaeI fragment are found on a 80 kb ClaI fragment. The ClaI sites at map positions 5 kb, 55 kb and 205 kb indicate the origin of these fragments. 4) A subclone of XTY26 has been generated which contains both Do33 and vector sequences. This places the vector sequences between Do33 and 2-34 completing a circle with the human DNA insert (Figure 1).

Most of the restriction endonucleases used to generate the pulsed-field gel map of XTY26 contain CpG dinucleotides in their recognition sequences. While this contributes to their underrepresentation in the genome, and therefore utility in long range restriction mapping, the methylation of mammalian DNA at these sites rendered a direct comparison of the XTY26 map to human chromosomal DNA all but useless. A fortunate exception was SfiI whose recognition sequence does not contain CpG and which generates a 120 kb SfiI fragment from XTY26 containing DSX293 and most of the DNA between this locus and Do33 (approximately 150 kb). The same 120 kb SfiI fragment was detected in human lymphocyte DNA from a normal individual confirming the integrity of at least a portion

of the human DNA sequence in XTY26. The integrity of the human insert is further supported by evidence from restriction maps of YACs in this area that show the probes 2-34 and Do33 markers to be approximately 210 kb apart. Consistent with these data is the finding that these markers are between 230 and 260 kb apart in XTY26.

XTY26 was deposited at the American Type Culture Collection, Rockville, MD, on December 5, 1990, under the terms of the Budapest Treaty, and has accession no. ATCC 74041.

Location of a DNA Probe Spanning Fragile X

To identify sequences which constitute the fragile site and to screen for DNA differences between normal and Fragile X individuals in the vicinity of the fragile site, sequences from XTY26 were used as hybridization probes. Localization of the fragile site was accomplished by first establishing a contig of λ-phage subclones between the two closest sequences which flanked the fragile site. A diagram of the relevant portion of XTY26 is shown in Figure 2.

The VK16 site (which had been utilized to isolate XTY26) has been localized proximal to the fragile site by *in situ* hybridization (Kremer, E. et al., *Am. J. Human Genet*. (1991) 49:656-661), incorporated herein by reference). Its position in XTY26 is shown in Figure 2. The distal end of the contig was established by initially screening the lambda library of XTY26 with an *Alu* PCR product (Nelson, D.L. et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6686), referred to as Alu2 (Figure 2). The subclone #91 was isolated with this probe and was subsequently shown by *in situ* hybridization to map distal to the fragile site. Riboprobes from each end of #91 were used to "walk away" from this locus and the direction of the "walk" was established by hybridization

back to blots of various restriction enzyme digests of XTY26. Each of the lambda subclones between #91 and VK16 was mapped relative to the fragile site by in situ hybridization.

The detailed steps of the above procedure are depicted in Figure 2, and are as follows. The letters at the beginning of each paragraph refer to the figure.

A: The "rare-cutter" restriction endonuclease map of the yeast artificial chromosome, XTY26, was determined by pulse-field gel electrophoresis (Kremer, E. et al., Am. J. Human Genet. (1991) 49:656-661). The locations of four probes (VK16, 2-34, Do33 and Alu 2) are indicated. Alu 2 was generated in a PCR using XTY26 DNA as template and the Alu consensus sequence oligo TC-65 (Nelson, D.L. et al., Proc. Natl. Acad. Sci. USA (1989) 86:6686) as primer. The localization of other probes has been reported previously (Kremer, E. et al., supra).

B and C: A contig of subcloned DNA fragments of XTY26 was generated by construction of a partial Sau3AI digest library in λGEM-3 (Promega), using the manufacturer's protocols and packaging extracts. The library was first screened with total human DNA, then the plaque-purified array of 108 clones was probed with Alu2 and VK16. Riboprobes were generated (again using the manufacturer's protocols and reagents) from the positive clones and used to "walk" towards and across the fragile site region. The direction of the "walk" was established by mapping these riboprobes back to the XTY26 restriction map. Each of the subclones was then used in fluorescence in situ hybridization to localize the fragile site with respect to the contig. This localization and its approximate boundaries are shown by dashed lines.

D and E: Each of the clones which flank and span the fragile site region, as defined by in situ hybridization, were used as probes on Southern blots of

s matic cell hybrid DNAs. These results confirmed the EcoRI restriction map across this region. The location of the breakpoints in hybrids Q1X and micro 21D are indicated by dashed lines.

F: This shows the restriction endonuclease map of the 5 kb EcoRI fragment which demonstrates instability in Fragile X individuals. The CpG region is indicated by the cluster of "rare-cutter" restriction endonuclease recognition sites.

G: Restriction fragments were used as hybridization probes to delineate the region of instability.

The *in situ* hybridization mapping delineated the sequences which appeared to "bridge" the fragile site to about 15 kb, although the extent and boundaries of this region could not be sharply defined. Each of the lambda clones which bridged the fragile site was then used as a hybridization probe against several somatic cell hybrid DNAs. Two of these, Q1X and micro 21D, had been constructed from a Fragile X parent cell line (Y75-1B-MI) in a way designed to break the X chromosome at the fragile site (Warren, S. et al., Proc. Natl. Acad. Sci. USA (1990) 87:3856). These hybrids have breakpoints which mapped within the same 5 kb EcoRI restriction fragment (Figures 2 and 3).

With respect to Figure 3, chromosomal DNA was isolated from the somatic hybrid cell line CY3, which contains the Xq26-qter region intact from a normal X chromosome (lane 1); Y75-1B-M1 (lane 2): Q1X (lane 3): Micro 21D (lane 4) and the mouse cell line A9, which is one parent line of CY3 (lane 5). The chromosomal DNA was subjected to cleavage with restriction endonuclease EcoRI, subjected to gel electrophoresis, and probed with nick-translated λ5. The Southern Blot obtained is shown in Figure 3. The kb EcoRI fragment normally expected,

which contains the Q1X and Micro 21D breakpoints and the Y75-1B-M1 instability, is arrowed in each lane. This is altered in mobility in Q1X, Micro 21D and Y75-1B-M1 as shown. The 5.3 and 1.3 kilobase EcoRI fragments flank the unstable fragment and are present in the Micro 21D and Q1X hybrids, respectively.

Cell line Y75-1B-M1 demonstrated an increase in size in the common breakpoint fragment from 5 to 5.9 kb. It appeared, therefore, that this variation might be associated with the fragile site, and this hypothesis was then tested.

The λ5 subclone containing the 5 kb EcoRI fragment was used as a probe on DNA from both normal and unrelated Fragile X Syndrome affected males. As depicted in Figure 4, DNA from four unrelated Fragile X Syndrome affected males (lanes 3 to 6) was digested with EcoRI and subjected to Southern blot analysis using subclone λ-#5 as probe. Comparison with normal male DNA (lane 1) and with a normal male from an affected pedigree (lane 2) revealed the altered mobility of the 5 kb EcoRI fragment to one or more high molecular weight bands in each of the affected individuals. Accordingly, it has been found that unrelated Fragile X Syndrome affected males demonstrate instability of DNA sequences at the site shown in Figure 2 as FRAXA.

No variation was observed between any normal individuals, whereas every Fragile X male showed an altered mobility of this sequence. The origin of this variability was localized further by using a series of restriction fragments from the 5 kb EcoRI fragment as probes. Fragments A, C and D (Figure 2G) all showed no variation between PstI digests of normal and affected individuals (data not shown). The 1.0 kb Pst fragment B was found to hybridize to repeat sequences in the human genome, whereas the 520 base pair fragment E (derived

from fragment B) hybridized strongly to a single PstI fragment which again demonstrated variation in size in unrelated Fragile X Syndrome affected individuals. Some Fragile X Syndrome individuals had from one to six recognizable bands of varying size and intensity. Others had multiple bands which manifested as a smear. In those males with only a smear, PCR amplification of the 520 bp band from their genomic DNA confirmed that this sequence was present and had not been deleted from their genomes (data not shown). The number of Fragile X genotype and normal DNA samples analyzed and the patterns of hybridization seen in them are summarized in Table 2. Abnormal bands were seen on Southern Blots EcoRI or PstI digests) in 61 Fragile X individuals from 18 families and 48 unrelated controls.

Table 2

		Single band of increased size	2-4 bands of increased size	Multiple bands of increased size ("smear")
Males :	Affected	5	5	11
:	Transmitting	3	1	1
Females:	Normal carriers	17	7	2
:	Affected	4	3	2
Normal Males	(n=26)	0	0	0
Females	(n=22)	0	0	0

Males were classifi d as affected by having expression of the fragile site and clinical features of the Fragile X Syndrome. Transmitting males were classified by their position in the pedigree or by having a high probability, on the basis of flanking DNA polymorphisms, of having the Fragile X genotype, and as normal by not having either fragile site expression or clinical features of the syndrome. Female carriers were classified as affected or normal on the basis of clinical features of the syndrome, regardless of fragile site expression.

The nature of this variable sequence was further investigated in Fragile X Syndrome pedigrees, as depicted in Figure 5. DNA from members of the illustrated Fragile X Syndrome pedigree was digested with PstI and subjected to Southern blot analysis using fragment E as probe. Pedigree symbols : unshaded, normal male (square) or female (circle); central dot, normal carrier male (square) or female (circle) not expressing Fragile X; half-shaded circle, normal female expressing Fragile X; shaded square, retarded fragile X syndrome male expressing Fragile X. Normal individuals in generation 3 had a less than 2% chance of carrying Fragile X based upon flanking DNA polymorphisms (Sutherland, G.R., and Mulley, J.C., <u>Clinical Genet.</u> (1990) 37:2-11).

This analysis demonstrated segregation of the variable sequence with the Fragile X genotype, with altered mobilities observed in nonpenetrant "transmitting" males and carrier females as well as affected males. The alteration in mobility varied within families where a single band was observed, and in the two families studied increased in size from generation to generation when transmitted by females, but not when

transmitted by males, and was larger in affected individuals than in normal carriers. The lack of a single hybridizing band in some Fragile X genotypes may reflect somatic heterogeneity occasionally leading to a smear, since the probe sequence is known to be present. Furthermore, in all cases where a band was observed, the variation was manifest as an increase in size, suggesting amplification or insertion. These properties suggest that the sequences inserted into or amplified from within the 1 kb PstI fragment are unstable in Fragile X individuals. The molecular basis for the instability is not clear because of difficulties in sequence analysis. However, the observation of repeat sequences within the unstable region suggests that the instability might be due to variation in the length or number of these repeats.

The restriction map of XTY26 which was derived from a Fragile X individual did not appear to differ from normal human DNA in the region of instability. This may be due to an undetected small difference in the size of the 1.0 kb PstI fragment or to deletion of the amplified region during cloning.

Nature of the Fragile X-Containing Fragment

The 1 kb Pst fragment is highly GC rich and in Fragile X affected individuals is refractory to PCR analysis. A high GC content is reflected in a CpG region which contains recognition sites for several CpG containing restriction enzymes. Three of these sites have been found to be subject to variations in methylation status, which segregates with Fragile X Syndrome phenotype but not genotype (Vincent, A. et al., Nature (1991) 349:624). The finding of sequences at the Fragile X locus which exhibit instability (presumably amplification or insertion), and which segregate with

genotype (regardless of fragile site expression or phenotype), suggests that the degree of size increase in these sequences might modulate fragile X expression and the associated syndrome. The immediate proximity of the unstable sequences to a CpG island, denoted $p(CCG)_n$, suggests interference with either the expression of a gene or the function of its product, as a molecular basis for the disease phenotype.

The sequence of the 1 kb PstI fragment is shown in Figure 6.

Utility of the Fragile X Probe

The previously mapped markers, Do33 and VK16, one distal and one proximal to the Fragile X locus, frame a 150 kb fragment which contains the fragile site as shown in Figure 1. Excision of this 150 kb fragment provides a more convenient probe than either of the closely associated markers. Further restriction and mapping of the 150 kb segment results in the preparation of a probe spanning the fragile site suitable for diagnosis.

The isolated 520 bp segment of the 1 kb Pst from the NheI site of the PstI set fragment forms a diagnostic reagent for direct detection of the Fragile X genotype. It will detect all Fragile X males by the altered mobility of a 1 kb PstI band or its apparent absence. It will, however, only reliably detect Fragile X females where there is a band or bands of altered size because, for those females where the abnormal band is a "smear," the pattern appears to be very similar to that of normal females. Testing Fragile X families with this probe can be used as a means of Fragile X phenotype prediction, as well as genotype identification.

The fragile site-containing probe is thus used for diagnosis (e.g., prenatal diagnosis or carrier

detection) by standard techn logy utilizing means to detect hybridization of the probe under appropriate stringency conditions to the abnormal sequence. Any suitable means for detection of hybridization can be used, including radioactive or fluorescent labeling of the probe. For effective use as a probe, a fragment of the 150 kb segment may be 10 to 10,000 nucleotides in length, preferably 50 to 1000 nucleotides in length, more preferably 100 1000 nucleotides in length. The probe may be prepared by enzymatic digestion of a larger fragment of DNA or may be synthesized.

Further, by altering the stringency of the conditions of hybridization the sequences corresponding to the Fragile X locus can be isolated from normal subjects, sequenced, and corresponding sequences used in genetic therapy to correct this defect. Thus, the present invention also provides a method to treat mental retardation caused by the presence of a Fragile X locus, which method comprises replacing, repairing or compensating for said fragile site DNA of the X chromosome of a subject with the corresponding fragile site DNA of a normal chromosome.

The availability of cloned sequences from the Fragile X locus also makes possible the identification of a protein product encoded by the cloned sequences. Such proteins may be identified by operably linking the cloned sequences to a promoter in an expression vector. Many appropriate expression vectors for this purpose are widely known in the art. See, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, 1990, Cold Spring Harbor Press, Cold Spring Harbor, NY. The protein product may be used for diagnostic or therapeutic purposes. Thus, for example, the presence, absence, or alteration of the protein product may correspond to the status of an affected individual. Similarly, the protein

product from a normal individual may be used to treat an affected individual with an altered protein product.

Furthermore, monoclonal or polyclonal antibodies against the protein product may be raised by a wide variety of techniques widely known in the art. These antibodies may be labeled and used in a variety of immunoassays, or, as described above, for therapeutic use in an affected individual. See, for example, Harlow, et al., Antibodies: A Laboratory Manual, 1988, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Isolation of Polymorphic Microsatellite AC Repeats (FRAXAC1 and FRAXAC2) Linked to Fragile X

The Southern blot hybridization using probes described above, while accurate in determining genotype, is a relatively slow procedure, particularly for prenatal diagnosis. Genotype can be determined just as accurately by linkage analysis where the fetus is unaffected and when informative markers show no recombination with the disease locus. When such markers are polymerase chain reaction (PCR)-based, then the affection status for at-risk pregnancies can be determined much more rapidly than with the Southern blot-based test. Therefore, characterization of AC repeat sequences in the immediate vicinity of the fragile X site $p(CCG)_n$ unstable element was undertaken as follows.

A. Identification of Microsatellite repeat sequence and design of PCR Printers

The 108 λ subclones of the yeast artificial chromosome XTY26 were screened in a random-primed reaction (Multiprime, Amersham) with synthetic poly(AC.GT) (Pharmacia) radioactively labeled with $\alpha^{32}P$-dCTP. AC repeat-containing DNA clones were identified by hybridization to this probe in 0.5 M sodium phosphate, pH

7.0, 7% SDS (without carrier DNA) at 65°C f r 16 hours and by washing at 65°C for 1 hour in 2x SSC.

DNA from positive clones was digested with either HaeIII, Sau3AI, HinPI, HpaII, RsaI, HinfI or TaqI, electrophoresed on 1.4% agarose gels, blotted onto nylon membranes (GeneScreen Plus, NEN-Dupont) and probed with ^{32}P-poly(AC.GT) as above. Digests which gave a hybridizing fragment of less than 600 base pairs were chosen for subcloning into M13mp18 for sequence analysis. The derived sequences were then used to design synthetic oligodeoxyribonucleotide primers suitable for PCR analysis of length variation in the AC repeat sequences. These sequences for PCR primers were chosen on the basis of their apparent uniqueness, their 50% GC composition and their lack of consecutive G residues which appear to interfere with chemical synthesis of oligodeoxyribonucleotides.

The markers from each microsatellite were subsequently termed FRAXAC1 (from λ12), FRAXAC2 (from λ25) and FRAXAC3 (from λ26).

B. <u>Heterozygosity of Microsatellite Regions</u>

These primers and microsatellite regions were used to determine genotypes as follows. PCR incubations were performed in 10 μl volumes in a Perkin Elmer-Cetus thermocycler for 10 cycles at 94°C for 60s, at 60°C for 90s and then 72°C for 90s, followed by 25 cycles at 94°C for 60s, at 55°C for 90s, and at 72°C for 90s. The volume was adjusted to 40 μl with formamide loading buffer (95% formamide, 1 mM EDTA, 0.01% bromophenyl blue, 0.01% xylene cyanol). After denaturation at 90°C for 3 minutes, 2.5 μl aliquots of each reaction mixture were subjected to electrophoresis in 6% polyacrylamide denaturing (7 M urea) gels. Genotypes were determined after autoradiography for 18 hours. Multipoint analysis

was based on genotypes of each marker in the 40 large kindred pedigrees of the Centre D'Etude du Polymorphisme Humain (CEPH) and was carried out using the LINKAGE (version 4.9) package for use with the CEPH three-generation families. The observed heterozygosities of FRAXAC3 in 18 unrelated females was only 16% and so the characterization of this marker was not pursued further. The observed heterozygosities of FRAXAC1 and FRAXAC2 were found to be 45% and 80%, respectively, in 40 unrelated females. However, none of the females homozygous for FRAXAC2 were heterozygous for FRAXAC1 and so the combined observed heterozygosity was also 80% (Table 3). This indicates linkage disequilibrium between the two markers.

Table 3

Marker	Allele $(AC)_n$	Allele Frequency	% Heterozygosity Observed	% Heterozygosity Expected*
FRAXAC1	19	0.0625	45	43.5
	18	0.0125		
	17	0.725		
	16	0.1875		
	15	0.0125		
FRAXAC2	23	0.009	80	71
	19	0.018		
	18	0.073		
	17	0.477		
	16	0.193		
	15	0.037		
	14	0.110		
	13	0.083		

*Based on observed allele frequencies.

C. <u>Genotyping of FRAXAC1 and FRAXAC2</u>

The genotypes of both markers were determined in the 40 unrelated families from CEPH. No recombination was observed between them.

Fragile X-affected pedigrees who had previously been shown to have recombinants in the vicinity of the fragile site were genotyped with FRAXAC2. Of those individuals who were informative, no recombination was found between this marker and the Fragile X genotype (as determined by hybridization with a

subclone of the PstI fragment described below) or Fragile X phenotype (as determined by Fragile X expression and/or intellectual handicap). Thus, these markers are considerably more closely linked to Xq27.3 than the previously mapped AC repeat sequences VK23AC and VK144AC. Analysis with FRAXAC1 was not undertaken because of the high degree of linkage disequilibrium between the two markers. The subclone used as probe denoted pfxa3 is the NheI to PstI subclone of the PstI band as shown in Figure 8.

D. Alternative Method for Fragile X Diagnosis

Thus, an alternative approach to rapid diagnostic analysis of Fragile X Syndrome would be to use these tightly linked, highly informative genetic markers. Together with the pfxa3 hybridization probe, these new FRAXA markers provide a rational approach to prenatal diagnosis in Fragile X pedigrees. This involves analysis of chorionic villi sample DNA (CVS) with the AC repeat markers FRAXA1 or FRAXA2.3 to haplotype the FRAXA locus, followed by the Southern blots with the pfxa3 as probe to detect amplification of the $p(CCG)_n$ repeat. The initial microsatellite results allow rapid determination of unaffected status in 40% of cases whereas the prediction of phenotype for individuals with the FRAXA genotype will be subsequently determined by the size of pfxa3 hybridizing fragments.

CLAIMS

1. A purified and isolated DNA molecule of not more than 275 kb which comprises the human Fragile X locus.

2. The DNA molecule of claim 1 which is XTY26.

3. A purified and isolated DNA molecule of not more than 150 kb which comprises the human Fragile X locus.

4. The DNA of claim 3 which comprises the segment of XTY26 located between VK16 and Do33.

5. A DNA probe suitable for diagnosis of Fragile X Syndrome which comprises an oligomeric DNA comprising the human Fragile X locus.

6. The DNA probe of claim 5 which comprises a 1 kb Pst fragment of XTY26.

7. The DNA probe of claim 6 which is pfxa.

8. The DNA probe of claim 6 which comprises the sequence or an effective fragment thereof:

10	20	30	40	50
CTGCAGAAAT	GGGCGTTCTG	GCCCTCGCGA	GGCAGTGCGA	CCTGTCACCG
60	70	80	90	100
CCCTTCAGCC	TTCCCGCCCT	CCACCAAGCC	CGCGCACGCC	CGGCCCGCGC
110	120	130	140	150
GTCTGTCTTT	CGACCCGGCA	CCCCGGCCGG	TTCCCAGCTG	CGCGCATGCC
160	170	180	190	200
GGCGCTCCCA	GGCCACTTGA	AGAGAGAGGG	CGGGGCCGAG	GGCTGAGCC

```
         210        220        230        240        250
  GCGGGGGGAG GGAACAGCGT TGATCACGTG ACGTGGTTTC AGTGTTTACA
         260        270        280        290        300
  CCCGCAGCGG GCCCGGGGGT TCGGCCTCAG TCAGGCGCTC AGCTCCGTTT
         310        320        330        340        350
  CGTTTCACTT CCGGTGGAGG GCCGCCTCTG AGCGGGCGGC GGGCCGACGG
         360        370        380        390        400
  CGAGCGCGGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC
         410        420        430        440        450
  GGCGGCGGTG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG
         460        470        480        490        500
  CGGCGGCGGC GGCGGCGGCG GCGGCCCGGA GCCACCTCTT CGGGGGCGGG
         510        520        530        540        550
  CTCCCGGCGC TAGCAGGGCT GAAGAGAAGA TGGAGGAGCT GGTGGTGGAA
         560        570        580        580        600
  CTGCGGGGCT CCAATGGCGC TTTCTACAAG GTACTTGGCT CTAGGGCAGG
         610        620        630        640        650
  CCCCATCTTC GCCCTTCCTT CCCTCCCTTT TCTTCTTGGT GTCGGCGGGA
         660        670        680        690        700
  GGCAGGCCCG GGGCCCTCTT CCCGAGCACC GCGCCTGGGT GCCAGGGCAC
         710        720        730        740        750
  GCTCGGCGGG ATGTTGTTGG AGGGAAGGAC TGGACTTGGG GCCTGTTGGA
         760        770        780        790        800
  AGCCCCTCTC CGACTCCGAG AGGCCCTAGC GCCTATCGAA ATGAGAGACC
         810        820        830        840        850
  AGCGAGGAGA GGGTTCTCTT TCGGCGCCGA GCCCGCCGGG GTGAGCTGGG
         860        870        880        890        900
  GATGGGCGAG GGCCGGCGGC AGGTACTAGA GCCGGGCGGG AAGGGCCGAA
         910        920        930        940        950
  ATCGGCGCTA AGTGACGGCG ATGGCTTATT CCCCCTTTCC TAAACATCAT
         960        970        980        990       1000
  CTCCCAGCGG GATCCGGGCC TGTCGTGTGG GTAGTTGTGG AGGAGCGGGG
```

```
      1010       1020       1030
GGCGCTTCAG CCGGGCCACC TCCTGCAG.
```

9. The probe of claim 6 conjugated to label.

10. The probe of claim 9 wherein the label is fluorescent or radioactive.

11. A purified and isolated DNA comprising the sequence:

```
        10         20         30         40         50
CTGCAGAAAT GGGCGTTCTG GCCCTCGCGA GGCAGTGCGA CCTGTCACCG
        60         70         80         90        100
CCCTTCAGCC TTCCCGCCCT CCACCAAGCC CGCGCACGCC CGGCCCGCGC
       110        120        130        140        150
GTCTGTCTTT CGACCCGGCA CCCCGGCCGG TTCCCAGCTG CGCGCATGCC
       160        170        180        190        200
GGCGCTCCCA GGCCACTTGA AGAGAGAGGG CGGGGCCGAG GGGCTGAGCC
       210        220        230        240        250
GCGGGGGGAG GGAACAGCGT TGATCACGTG ACGTGGTTTC AGTGTTTACA
       260        270        280        290        300
CCCGCAGCGG GCCCGGGGGT TCGGCCTCAG TCAGGCGCTC AGCTCCGTTT
       310        320        330        340        350
CGTTTCACTT CCGGTGGAGG GCCGCCTCTG AGCGGGCGGC GGGCCGACGG
       360        370        380        390        400
CGAGCGCGGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC
       410        420        430        440        450
GGCGGCGGTG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG
       460        470        480        490        500
CGGCGGCGGC GGCGGCGGCG GCGGCCCGGA GCCACCTCTT CGGGGGCGGG
       510        520        530        540        550
CTCCCGGCGC TAGCAGGGCT GAAGAGAAGA TGGAGGAGCT GGTGGTGGAA
       560        570        580        580        600
CTGCGGGGCT CCAATGGCGC TTTCTACAAG GTACTTGGCT CTAGGGCAGG
```

```
       610        620        630        640        650
CCCCATCTTC GCCCTTCCTT CCCTCCCTTT TCTTCTTGGT GTCGGCGGGA
       660        670        680        690        700
GGCAGGCCCG GGGCCCTCTT CCCGAGCACC GCGCCTGGGT GCCAGGGCAC
       710        720        730        740        750
GCTCGGCGGG ATGTTGTTGG AGGGAAGGAC TGGACTTGGG GCCTGTTGGA
       760        770        780        790        800
AGCCCCTCTC CGACTCCGAG AGGCCCTAGC GCCTATCGAA ATGAGAGACC
       810        820        830        840        850
AGCGAGGAGA GGGTTCTCTT TCGGCGCCGA GCCCGCCGGG GTGAGCTGGG
       860        870        880        890        900
GATGGGCGAG GGCCGGCGGC AGGTACTAGA GCCGGGCGGG AAGGGCCGAA
       910        920        930        940        950
ATCGGCGCTA AGTGACGGCG ATGGCTTATT CCCCCTTTCC TAAACATCAT
       960        970        980        990       1000
CTCCCAGCGG GATCCGGGCC TGTCGTGTGG GTAGTTGTGG AGGAGCGGGG
      1010       1020       1030
GGCGCTTCAG CCGGGCCACC TCCTGCAG.
```

12. A purified and isolated DNA molecule of less than 50 nucleotides flanking a polymorphic microsatellite repeat sequence within the DNA molecule of claim 1, wherein said DNA molecule is a primer for amplification of said microsatellite repeat sequence.

13. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence

GATCTAATCA ACATCTATAG ACTTTATT.

14. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence

AGGCTTGGA GTGCAGTGGG CAATCT.

15. The DNA molecule of claim 12, wherein said microsatellite repeat s quence comprises the nucleotide sequence

$(GT)_n$CAGTCTCA CTCTGTCACTC,

wherein n is 1-100.

16. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence

GACTGCTCCGG AAGTTGAATC CTCA.

17. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence

AGACAGGATC TCACTCTGTC ACCTAG.

18. The DNA molecule of claim 12, wherein said microsatellite repeat sequence comprises the nucleotide sequence

GTATTT TTGCAAAGTT TGTCTTTCAG TATTTTATTT$(GT)_n$ ATATATATAT ATTTTTTTTT TTTTTTTTAA,

wherein n is 1-100.

19. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence

GTACTGTATC AGTTATAACC CTATG.

20. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence

CAAAT TGAAGGTTTG TGGAAACCT.

21. The DNA molecule of claim 12, wherein said microsatellite repeat sequence comprises the nucleotide sequence

TGTGT GTGTGC$(GT)_n$ATGCAT ACCCAAGACT TATCTTATAC AGGTATGCCT TGTTTTATTG CACTTTGCAA ATACTGCATT TTTTT,

wherein n is 1-100.

22. A purified protein product encoded by the DNA of claim 1.

23. A monoclonal antibody immunologically reactive with the protein of claim 22.

24. A polyclonal antibody immunologically reactive with the protein of claim 22.

25. A method to diagnose the presence or absence of a Fragile X allele in a subject which method comprises digesting the DNA of said subject to obtain DNA fragments, separating the fragments according to size, and detecting a fragment spanning a DNA sequence comprising said allele among the fragments by hybridization to the probe of claim 6.

26. A method to treat mental retardation caused by the presence of a Fragile X allele, which method comprises replacing, repairing or compensating for DNA spanning a DNA sequence comprising said allele of a subject with the corresponding DNA sequence of a normal chromosome.

27. A method to treat mental retardation caused by the presence of a Fragile X allele, which method comprises replacing or compensating for protein encoded by the DNA sequence comprising said allele of a subject with the corresponding protein of a normal chromosome.

28. A method for screening for a Fragile X allele in a subject, comprising

a) designating a microsatellite repeat allele closely linked to Xq27.3 in a family, wherein said subject is a member of the family; and

b) screening said subject for presence of the microsatellite repeat allele closely linked to Xq27.3.

29. The method of claim 28, wherein step (b) further comprises amplification of said microsatellite repeat sequence with the primer of claim 12.

30. The method of claim 28, further comprising detecting amplification of a $p(CCG)_n$ repeat at said Xq27.3.

31. The method of claim 28, wherein detection of said repeat further comprises hybridization with the probe of claim 7.

32. The method of claim 28, wherein said microsatellite repeat allele is closely linked to Xq27.3 in a population.

AMENDED CLAIMS

[received by the International Bureau on 24 June 1992 (24.06.92); original claims 8,11,13-21 amended; other claims unchanged (7 pages)]

1. A purified and isolated DNA molecule of not more than 275 kb which comprises the human Fragile X locus.

2. The DNA molecule of claim 1 which is XTY26.

3. A purified and isolated DNA molecule of not more than 150 kb which comprises the human Fragile X locus.

4. The DNA of claim 3 which comprises the segment of XTY26 located between VK16 and Do33.

5. A DNA probe suitable for diagnosis of Fragile X Syndrome which comprises an oligomeric DNA comprising the human Fragile X locus.

6. The DNA probe of claim 5 which comprises a 1 kb Pst fragment of XTY26.

7. The DNA probe of claim 6 which is pfxa.

8. The DNA probe of claim 6 which comprises the sequence (SEQ ID NO 1) or an effective fragment thereof:

10	20	30	40	50
CTGCAGAAAT	GGGCGTTCTG	GCCCTCGCGA	GGCAGTGCGA	CCTGTCACCG
60	70	80	90	100
CCCTTCAGCC	TTCCCGCCCT	CCACCAAGCC	CGCGCACGCC	CGGCCCGCGC
110	120	130	140	150
GTCTGTCTTT	CGACCCGGCA	CCCCGGCCGG	TTTCCCAGCT	CGCGCATGCC
160	170	180	190	200
GGCGCTCCCA	GGCCACTTGA	AGAGAGAGGG	CGGGGCCGAG	GGGCTGAGCC
210	220	230	240	250
GCGGGGGGAG	GGAACAGCGT	TGATCACGTG	ACGTGGTTTC	AGTGTTTACA
260	270	280	290	300
CCCGCAGCGG	GCCCGGGGGT	TCGGCCTCAG	TCAGGCGCTC	AGCTCCGTTT
310	320	330	340	350

CGTTTCACTT CCGGTGGAGG GCCGCCTCTG AGCGGGCGGC GGGCCGACGG
360 370 380 390 400
CGAGCGCGGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC
410 420 430 440 450
GGCGGCGGTG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG
460 470 480 490 500
CGGCGGCGGC GGCGGCGGCG GCGGCCCGGA GCCACCTCTT CGGGGGCGGG
510 520 530 540 550
CTCCCGGCGC TAGCAGGGCT GAAGAGAAGA TGGAGGAGCT GGTGGTGGAA
560 570 580 590 600
CTGCGGGGCT CCAATGGCGC TTTCTACAAG GTACTTGGCT CTAGGGCAGG
610 620 630 640 650
CCCCATCTTC GCCCTTCCTT CCCTCCCTTT TCTTCTTGGT GTCGGCGGGA
660 670 680 690 700
GGCAGGCCCG GGGCCCTCTT CCCGAGCACC GCGCCTGGGT GCCAGGGCAC
710 720 730 740 750
GCTCGGCGGG ATGTTGTTGG AGGGAAGGAC TGGACTTGGG GCCTGTTGGA
760 770 780 790 800
AGCCCCTCTC CGACTCCGAG AGGCCCTAGC GCCTATCGAA ATGAGAGACC
810 820 830 840 850
AGCGAGGAGA GGGTTCTCTT TCGGCGCCGA GCCCGCCGGG GTGAGCTGGG
860 870 880 890 900
GATGGGCGAG GGCCGGCGGC AGGTACTAGA GCCGGGCGGG AAGGGCCGAA
910 920 930 940 950
ATCGGCGCTA AGTGACGGCG ATGGCTTATT CCCCCTTTCC TAAACATCAT
960 970 980 990 1000
CTCCCAGCGG GATCCGGGCC TGTCGTGTGG GTAGTTGTGG AGGAGCGGGG

```
      1010       1020       1030
GGCGCTTCAG CCGGGCCACC TCCTGCAG.
```

9. The probe of claim 6 conjugated to label.

10. The probe of claim 9 wherein the label is fluorescent or radioactive.

11. A purified and isolated DNA comprising the sequence (SEQ ID NO 1):

```
        10         20         30         40         50
CTGCAGAAAT GGGCGTTCTG GCCCTCGCGA GGCAGTGCGA CCTGTCACCG
        60         70         80         90        100
CCCTTCAGCC TTCCCCGCCC CCACCAAGCC CGCGCACGCC CGGCCCGCGC
       110        120        130        140        150
GTCTGTCTTT CGACCCGGCA CCCCGGCCGG TTCCCAGCTG CGCGCATGCC
       160        170        180        190        200
GGCGCTCCCA GGCCACTTGA AGAGAGAGGG CGGGGCCGAG GGGCTGAGCC
       210        220        230        240        250
GCGGGGGGAG GGAACAGCGT TGATCACGTG ACGTGGTTTC AGTGTTTACA
       260        270        280        290        300
CCCGCAGCGG GCCCGGGGGT TCGGCCTCAG TCAGGCGCTC AGCTCCGTTT
       310        320        330        340        350
CGTTTCACTT CCGGTGGAGG GCCGCCTCTG AGCGGGCGGC GGGCCGACGG
       360        370        380        390        400
CGAGCGCGGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC
       410        420        430        440        450
GGCGGCGGTG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG
       460        470        480        490        500
CGGCGGCGGC GGCGGCGGCG GCGGCCCGGA GCCACCTCTT CGGGGGCGGG
       510        520        530        540        550
CTCCCGGCGC TAGCAGGGCT GAAGAGAAGA TGGAGGAGCT GGTGGTGGAA
       560        570        580        590        600
CTGCGGGGCT CCAATGGCGC TTTCTACAAG GTACTTGGCT CTAGGGCAGG
```

610	620	630	640	650
CCCCATCTTC	GCCCTTCCTT	CCCTCCCTTT	TCTTCTTGGT	GTCGGCGGGA
660	670	680	690	700
GGCAGGCCCG	GGGCCCTCTT	CCCGAGCACC	GCGCCTGGGT	GCCAGGGCAC
710	720	730	740	750
GCTCGGCGGG	ATGTTGTTGG	AGGGAAGGAC	TGGACTTGGG	GCCTGTTGGA
760	770	780	790	800
AGCCCCTCTC	CGACTCCGAG	AGGCCCTAGC	GCCTATCGAA	ATGAGAGACC
810	820	830	840	850
AGCGAGGAGA	GGGTTCTCTT	TCGGCGCCGA	GCCCGCCGGG	GTGAGCTGGG
860	870	880	890	900
GATGGGCGAG	GGCCGGCGGC	AGGTACTAGA	GCCGGGCGGG	AAGGGCCGAA
910	920	930	940	950
ATCGGCGCTA	AGTGACGGCG	ATGGCTTATT	CCCCCTTTCC	TAAACATCAT
960	970	980	990	1000
CTCCCAGCGG	GATCCGGGCC	TGTCGTGTGG	GTAGTTGTGG	AGGAGCGGGG
1010	1020	1030		
GGCGCTTCAG	CCGGGCCACC	TCCTGCAG.		

12. A purified and isolated DNA molecule of less than 50 nucleotides flanking a polymorphic microsatellite repeat sequence within the DNA molecule of claim 1, wherein said DNA molecule is a primer for amplification of said microsatellite repeat sequence.

13. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence (SEQ ID NO 2)

GATCTAATCA ACATCTATAG ACTTTATT.

14. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence (SEQ ID NO 3)

AGGCTTGGA GTGCAGTGGG CAATCT.

15. The DNA molecule of claim 12, wherein said microsatellite repeat sequence comprises the nucleotide sequence (SEQ ID NO 4)

$(GT)_n$CAGTCTCA CTCTGTCACTC,

wherein n is 1-100.

16. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence (SEQ ID NO 5)

GACTGCTCCGG AAGTTGAATC CTCA.

17. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence (SEQ ID NO 6)

AGACAGGATC TCACTCTGTC ACCTAG.

18. The DNA molecule of claim 12, wherein said microsatellite repeat sequence comprises the nucleotide sequence (SEQ ID NO 7)

GTATTT TTGCAAAGTT TGTCTTTCAG TATTTTATTT$(GT)_n$ ATATATATAT ATTTTTTTTT TTTTTTTTAA,

wherein n is 1-100.

19. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence (SEQ ID NO 8)

GTACTGTATC AGTTATAACC CTATG.

20. The DNA molecule of claim 12, wherein said DNA molecule comprises the nucleotide sequence (SEQ ID NO 9)

CAAAT TGAAGGTTTG TGGAAACCT.

21. The DNA molecule of claim 12, wherein said microsatellite repeat sequence comprises the nucleotide sequence (SEQ ID NO 10)

TGTGT GTGTGC$(GT)_n$ATGCAT ACCCAAGACT TATCTTATAC AGGTATGCCT TGTTTTATTG CACTTTGCAA ATACTGCATT TTTTT,

wherein n is 1-100.

22. A purified protein product encoded by the DNA of claim 1.

23. A monoclonal antibody immunologically reactive with the protein of claim 22.

24. A polyclonal antibody immunologically reactive with the protein of claim 22.

25. A method to diagnose the presence or absence of a Fragile X allele in a subject which method comprises digesting the DNA of said subject to obtain DNA fragments, separating the fragments according to size, and detecting a fragment spanning a DNA sequence comprising said allele among the fragments by hybridization to the probe of claim 6.

26. A method to treat mental retardation caused by the presence of a Fragile X allele, which method comprises replacing, repairing or compensating for DNA spanning a DNA sequence comprising said allele of a subject with the corresponding DNA sequence of a normal chromosome.

27. A method to treat mental retardation caused by the presence of a Fragile X allele, which method comprises replacing or compensating for protein encoded by the DNA sequence comprising said allele of a subject with the corresponding protein of a normal chromosome.

28. A method for screening for a Fragile X allele in a subject, comprising

a) designating a microsatellite repeat allele closely linked to Xq27.3 in a family, wherein said subject is a member of the family; and

b) screening said subject for presence of the microsatellite repeat allele closely linked to Xq27.3.

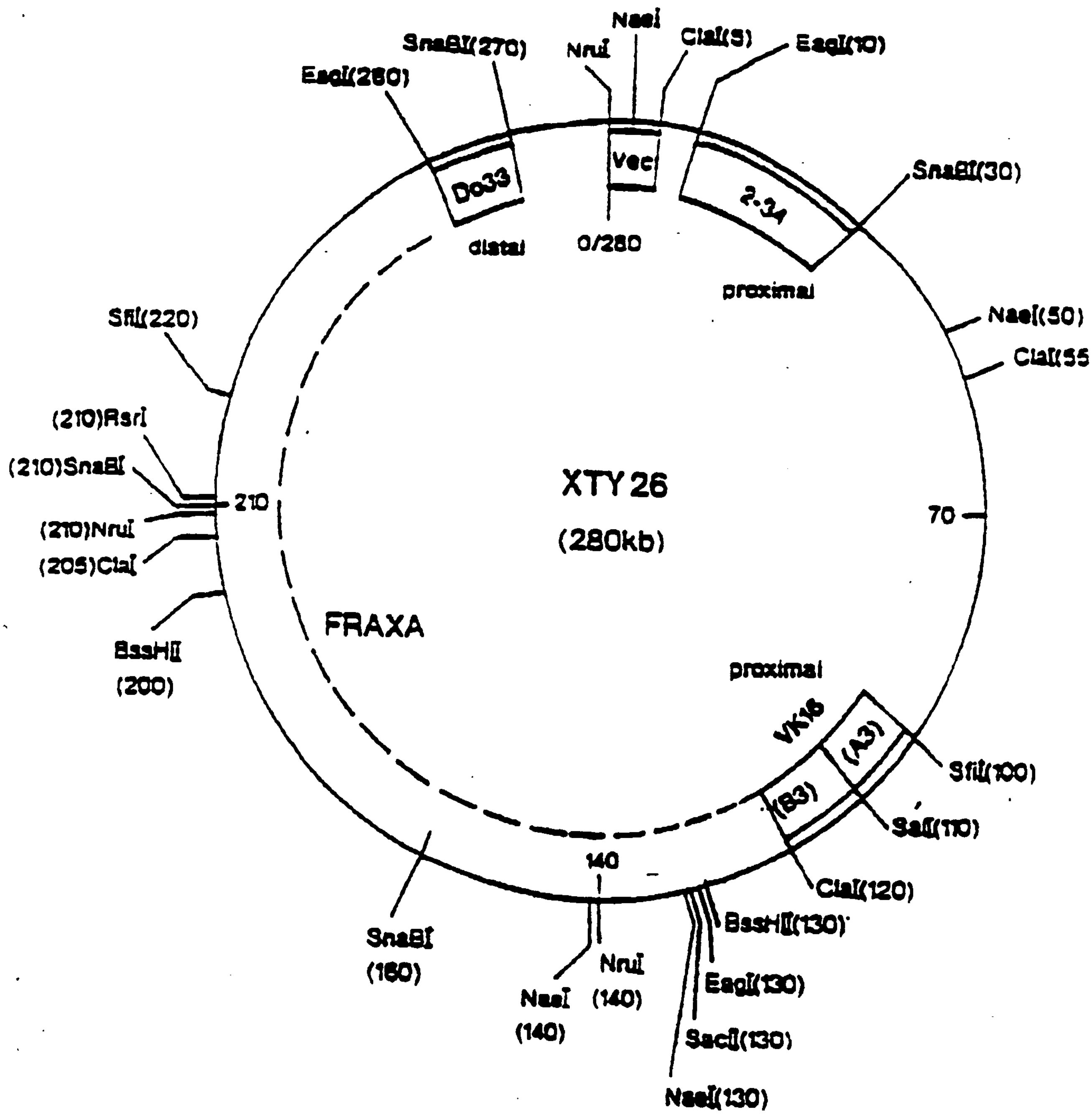
29. The method of claim 28, wherein step (b) further comprises amplification of said microsatellite repeat sequence with the primer of claim 12.

30. The method of claim 28, further comprising detecting amplification of a $p(CCG)_n$ repeat at said Xq27.3.

31. The method of claim 28, wherein detection of said repeat further comprises hybridization with the probe of claim 7.

32. The method of claim 28, wherein said microsatellite repeat allele is closely linked to Xq27.3 in a population.

FIGURE 1


NaeI
NruI
ClaI(5)
EagI(10)
SnaBI(270)
EagI(280)
Vec
Do33
2-34
SnaBI(30)
distal
0/280
proximal
NaeI(50)
ClaI(55)
SfiI(220)
(210)RsrI
(210)SnaBI
210
(210)NruI
(205)ClaI
XTY 26
(280kb)
70
BssHII
(200)
FRAXA
proximal
VK16
(A3)
(B3)
SfiI(100)
SalI(110)
ClaI(120)
140
BssHII(130)
SnaBI
(160)
NruI
(140)
NaeI
(140)
EagI(130)
SacII(130)
NaeI(130)

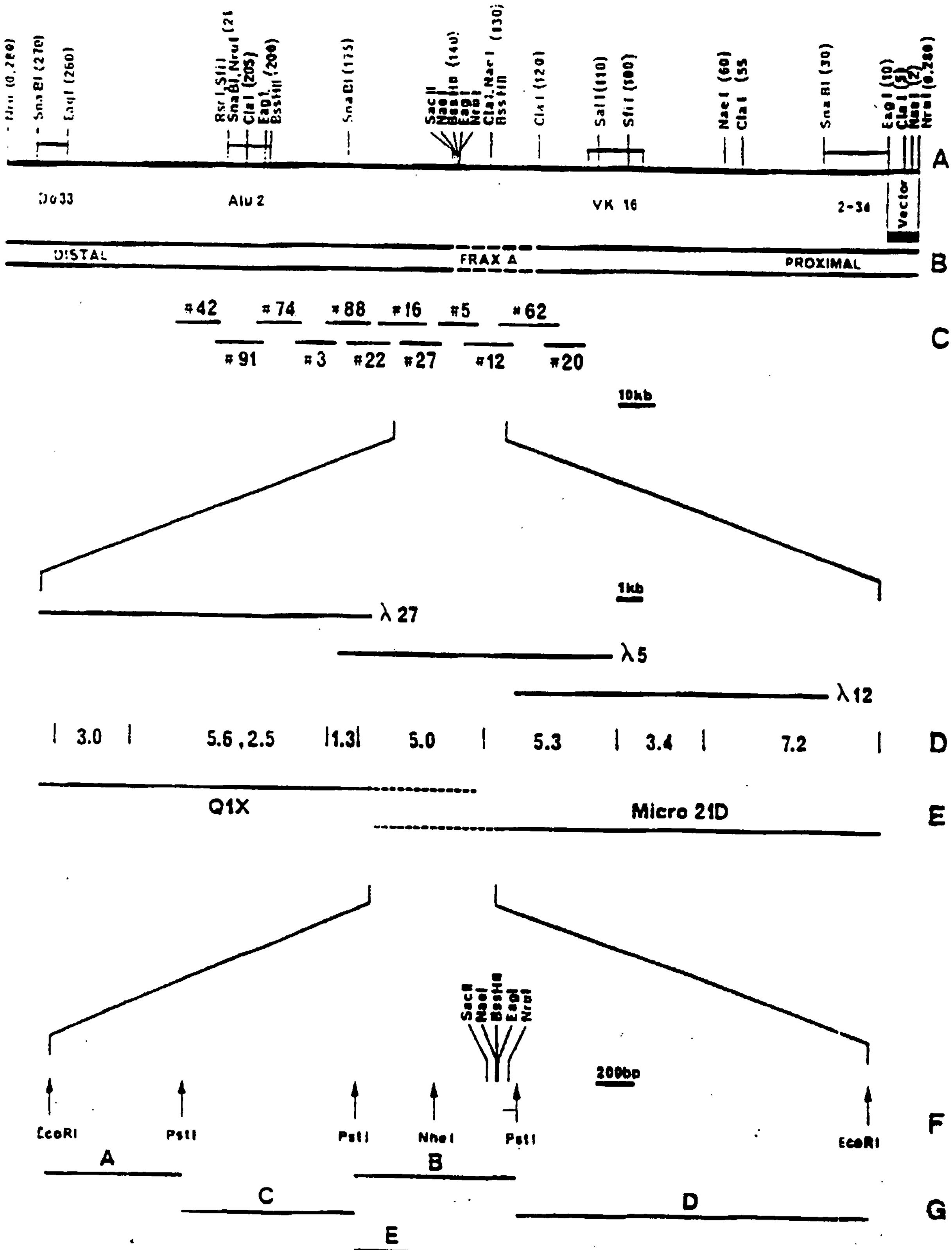

A
B
C
D
E
F
G
DISTAL
FRAX A
PROXIMAL
Alu 2
VK 16
2-34
Vector
#42
#74
#88
#16
#5
#62
#91
#3
#22
#27
#12
#20
10kb
1kb
λ27
λ5
λ12
3.0
5.6, 2.5
1.3
5.0
5.3
3.4
7.2
Q1X
Micro 21D
SacII
NaeI
BssHII
EagI
NruI
200bp
EcoRI
PstI
NheI
PstI
EcoRI
A
B
C
D
E

Figure 2

FIGURE 3

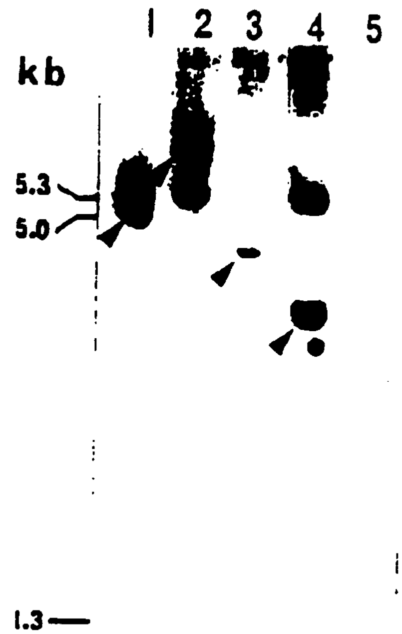


FIGURE 4

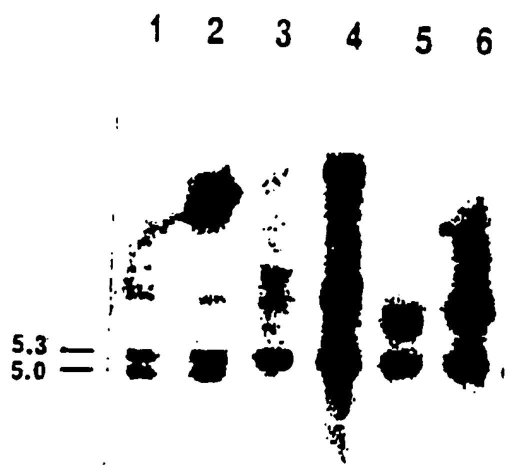


FIGURE 5

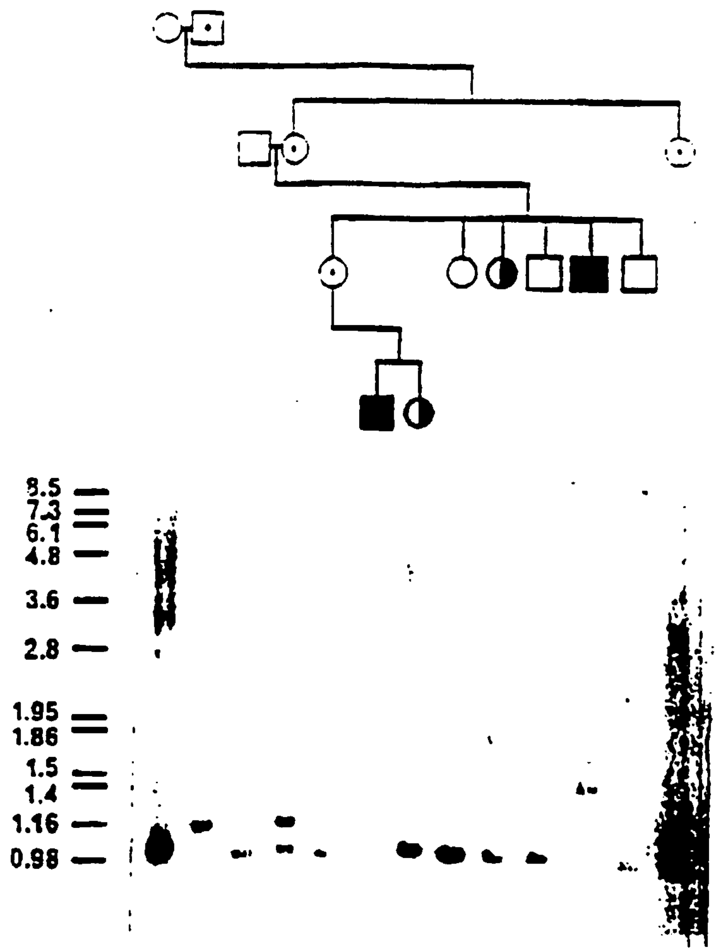


Figure 6

10 20 30 40 50 60
CTGCAGAAAT GGGCGTTCTG GCCCTCGCGA GGCAGTGCGA CCTGTCACCG CCCTTCAGCC

70 80 90 100 110 120
TTCCCGCCCT CCACCAAGCC CGCGCACGCC CGGCCCGCGC GTCTGTCTTT CGACCCGGCA

130 140 150 160 170 180
CCCCGGCCGG TTCCCAGCTG CGCGCATGCC GGCGCTCCCA GGCCACTTGA AGAGAGAGGG

190 200 210 220 230 240
CGGGGCCGAG GGGCTGAGCC TCGCGGGGAG GGAACAGCGT TGATCACGTG ACGTGGTTTC

250 260 270 280 290 300
AGTGTTTACA CCCGCAGCGG GCCCGGGGGT TCGGCCTCAG TCAGGCGCTC AGCTCCGTTT

310 320 330 340 350 360
CGTTTCACTT CCGGTGGAGG GCCGCCTCTG AGCGGGCGGC GGGCCGACGG CGAGCGCGGG

370 380 390 400 410 420
CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC GGCGGCGGTG GCGGCGGCGG

430 440 450 460 470 480
CGGCGGCGGC GGCGGCGGCG GCGGCGGCGG CGGCGGCGGC GGCGGCGGCG GCGGCCCGGA

490 500 510 520 530 540
GCCACCTCTT CGGGGGCGGG CTCCCGGCGC TAGCAGGGCT GAAGAGAAGA TGGAGGAGCT

550 560 570 580 590 600
GGTGGTGGAA GTGCGGGGCT CCAATGGCGC TTTCTACAAG GTACTTGGCT CTAGGGCAGG

610 620 630 640 650 660
CCCCATCTTC GCCCTTCCTT CCCTCCCTTT TCTTCTTGGT GTCGGCGGGA GGCAGGCCCG

670 680 690 700 710 720
GGGCCCTCTT CCCGAGCACC GCGCCTGGGT GCCAGGGCAC GCTCGGCGGG ATGTTGTTGG

730 740 750 760 770 780
AGGGAAGGAC TGGACTTGGG GCCTGTTGGA AGCCCCTCTC CGACTCCGAG AGGCCCTAGC

790 800 810 820 830 840
GCCTATCGAA ATGAGAGACC AGCGAGGAGA GGGTTCTCTT TCGGCGCCGA GCCCGCCGGG

850 860 870 880 890 900
GTGAGCTGGG GATGGGCGAG GGCCGGCGGC AGGTACTAGA GCCGGGCGGG AAGGGCCGAA

910 920 930 940 950 960
ATCGGCGCTA AGTGACGGCG ATGGCTTATT CCCCCTTTCC TAAACATCAT CTCCCAGCGG

970 980 990 1000 1010 1020
GATCCGGGCC TGTCGTGTGG GTAGTTGTGG AGGAGCGGGG GGCGCTTCAG CCGGGCCACC

1030 1040 1050 1060 1070 1080
TCCTGCAG

Figure 7

a. GATCTAATCA ACATCTATAG ACTTTATTGT GTGTGTGTGT GTGTGTGTGT GTATGTGTGT

GTCAGTCTCA CTCTGTCACT CAGGCTTGGA GTGCAGTGGC CAATCTCTGC TCACTGCAAC
100

CTCGCCTCCC AGCTTCAAGT GACTCTCATC ATGCCTCAGC CTCCTGAGTA GCTGGGATTA

CAGGCATGCA CCACCACACC CAGCTAATTT TTTGCATTTT TAGTAGAGTC GGCATTTCAC
200

TATGTTGGCC AGGCTGGTCT CGAACTTCTG GCCTCAAGTG ATC

b. GGCCCTAATC AGATTTCCAC AAATTCTGAC TTAATATTTG CCCGCTTATA TAACAGCTCT

TCTTTAACAA AAACAAGTAC TTTTCTCAAT AGAATTTTAC TAAGAAAGCT CTTTAGTAAA

ACATCGACAT TATACATACA ACATATCTCA GTATCTGCTG ATGAAGAACA CCAAAAAGAA

CCCAGATGTG ACTGCTCCGG AAGTTGAATC CTCAGTATTT TTGCAAAGTT TGTCTTTCAG

TATTTTATTT GTGTGTGTGT GTGTGTGTGT GTGTGTGTCT ATATATATAT ATTTTTTTTT
100

TTTTTTTTAA AGACAGGATC TCACTCTGTC ACCTAGGCTG GAGTGCAGTG CATGATCATG

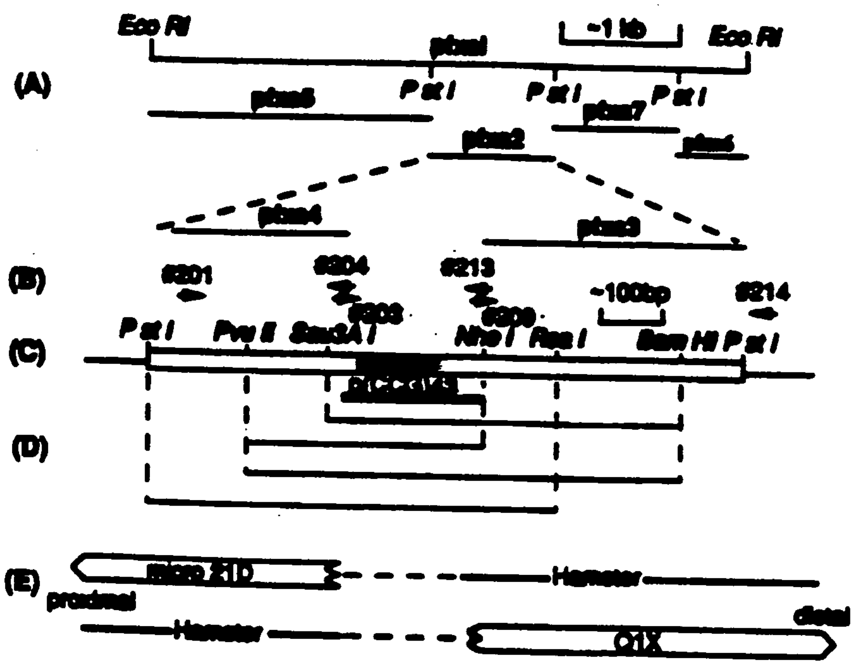
c. GTACTGTATC AGTTATAACC CTATGTGTGT GTGTGCGTGT GTGTGTGTGT GTGTATGCAT

ACCCAAGACT TATCTTATAC AGGTATGCCT TGTTTTATTG CACTTTGCAA ATACTGCATT
100

TTTTTCAAAT TGAAGGTTTG TGGAAACCTT TTTTTTGAGC AATTCTGTAG TGCCATTTTT

TTCAACGGCA TGTGTAC

Figure 8


Eco RI
~1 kb
Eco RI
(A)
Pst I
Pst I
Pst I
(B)
#201
#204
#213
~100bp
#214
(C)
Pst I
Pvu II
Sau3A I
Nhe I
Bam HI
Pst I
(D)
(E)
Hamster
proximal
Hamster
distal
Q1X

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00020

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): Please See Attached Sheet.
US CL : 436/6; 536/27; 530/350, 387; 424/95

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	436/6; 536/27; 530/350, 387; 424/95

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

CAS ONLINE, Medline, APS
search terms "fragile x", sequenc?, gene?

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
A	Ann. Rev. Genet., volume 20, issued 1986, Nussbaum et al. "Fragile X syndrome: a unique sequence in man," pages 109-145, see introduction pages 109-110	1-32
Y, P	Genomics, volume 9, issued January 1991, Suthers et al., "Genetic mapping of new RFLPs at Xq27-q28", pages 37-43, see abstract.	1-32
X/Y, P	Science, volume 251, issued 08 March 1991, Heitz et al."Isolation of Sequences That Span the Fragile X and Identification of a Fragile X related CpG Island", pages 1236-39, see abstract, page 1237.	5-9, 11/1-4, 10, 12-32

* Special categories of cited documents:[15]
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
21 APRIL 1992	04 MAY 1992
International Searching Authority[1]	Signature of Authorized Officer[20]
ISA/US	Scott A. Chambers

Form PCT/ISA/210 (second sheet)(May 1986)

FURTHER INF RMATION CONTINUED FROM PREVIOUS SHEETS

I. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

C12Q 1/68; C07H 15/12; C07K 3/00, 13/00; A61K 35/14; A01N 63/02; A61K 35/12